

**Autonomous weapons systems:
an analysis from human rights, humanitarian and
ethical artificial intelligence perspectives**

This paper is based on Wanda Muñoz' intervention at the briefing for diplomats on "The Normative and Operational Framework on Autonomous Weapon Systems" held on 28/05/2021.

Today I will share with you two angles of analysis on the issue of autonomous weapons. Firstly, I will share some perspectives of human rights activists and humanitarian workers. Secondly, I will talk about how the discussion on autonomous weapons could be informed by international discussions on ethical approaches to artificial intelligence (AI).

A/ Humanitarian and Human Rights Concerns

The topic of autonomous weapons should not be seen only as a military and diplomatic issue, because it is much more than that; it is also a social and humanitarian issue.

Unfortunately, what we see in the world today is that thousands of civilians have to deal with the consequences of violence and weapons daily, and that humanitarian organizations are at the forefront of the response.

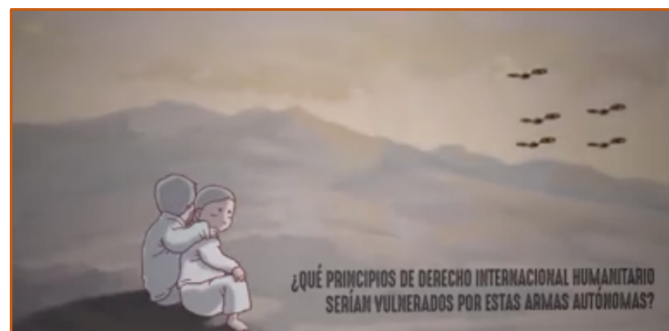
From this perspective, I would like to highlight four key reasons why autonomous weapon should be banned.

1. **Human dignity.** We believe that no one ever should face any risk of harm or death by autonomous weapons, because this would go against human dignity. This has already been said but we will keep insisting on it until autonomous weapons that target humans are banned.

What message do you think we would send to those populations where autonomous weapons would be deployed? In my view, we would be telling them: *"To us, your value as a human being is so little, that we don't care about the risks, and even a machine can do the job of killing you"*.

This is unacceptable. We do not want to live in a world that sends that message to anyone and neither do thousands of AI experts,¹ the UN Secretary General,² or the many countries that have already called for a ban on autonomous weapons systems.³

2. **Those of us who work with humanitarian organizations, particularly on victim assistance,** see the long-term consequences of weapons every day. We already face enough challenges as it is, without adding into the mix autonomous weapons with the range of challenges that they raise.⁴ I would also like us to think about the potential impact of these weapons on mental health: imagine if civilians -including children -already traumatized by conflict, suddenly see their lives **ALSO** threatened by autonomous weapons or a by a swarm of autonomous robots.⁵ That would certainly have a psychological impact that we cannot even measure, and I truly hope we never have to.



3. **The risk of error** which cannot be underestimated. I imagine we have all heard about the problems resulting from the use of algorithms for decision-making related to housing, employment, justice, health services and even COVID vaccination;⁶ or about the human rights violations that have resulted from the use of facial recognition for policing in different countries – to the point that Amazon,⁷ IBM⁸ and Microsoft⁹ actually established moratoriums on the use of this technology by police forces.

Because time and again, marginalized communities disproportionately face the negative impact of these technologies. For instance, studies by the MIT Media Lab led by scientists Joy Buolamwini, Timnit Gebru and Deborah Raji found that facial recognition has an error rate of only 0,8% when identifying white men; 19% when identifying dark-skinned men; and a staggering 34,7% when identifying dark-skinned women.¹⁰

4. **The risk of proliferation.** Considering the state of the world today, if we allow the development of fully autonomous weapons, countries already in situations of conflict or armed violence would probably be the ones where these weapons would be tested and deployed, which is already bad enough.

But eventually, none of us would be safe from them, because even if they are initially developed for use in conflict, they would probably find their way to countries to national police institutions, raising human rights concerns, and possibly to illegal and non-state armed groups. This assessment is not science fiction, it based on what happens with conventional weapons today.

So autonomous weapons are a risk in any context, from the perspectives of both, International Humanitarian Law and International Human Rights Law.

We believe these are just some of the reasons why autonomous weapons should be regulated; and those that cannot be used with meaningful human control or those that target humans should be banned.

B/ Perspectives from forums and frameworks on the ethics of artificial intelligence

Now, let me move on to the second topic: the link between our dialogues on autonomous weapons, and what is being discussed in forums related to the **ethics of artificial intelligence and emerging technologies**, because these issues are interrelated.

Many of the countries present here today actively participate, and some of them are even leading, **multistakeholder initiatives that aim to ensure that those technologies are human-centric and used responsibly**.¹¹ These have been going on for many years now and have input from policymakers, diplomats, human rights experts, scientists, and intergovernmental organizations.

So, I think it is useful to take a step back for a couple of minutes and to look at two important elements:

- a) what the **risks of autonomy** are in other sectors, and
- b) how these **risks are being addressed**, including through regulation.

Let me give you some examples:¹²

- Firstly, the negotiations of the **Recommendations on the Ethics of Artificial Intelligence** led by UNESCO, which specifically warns against the “*new types of ethical issues that include... weaponization*” and require countries to examine to what extent this technology poses a risk in terms of “*transgression of safety and human rights*”.¹³ Would we not say that autonomous weapons pose an enormous risk of transgression of human rights?

- Secondly, a resolution by the **African Commission on Human and Peoples' Rights**, which specifically -and I cite- *“Appeals to State Parties to ensure that all AI technologies, robotics and other new and emerging technologies which have far reaching consequences for humans must remain under **meaningful human control** to ensure that the threat that they pose to fundamental human rights is averted. The emerging norm of maintaining meaningful human control over AI technologies, robotics and other new and emerging technologies **should be codified as a human rights principle.**”*¹⁴ This is, of course, extremely relevant to our discussion.
- Thirdly, the **OECD Ethical principles for AI**, which have been adhered to by all OECD member countries and by six non-member countries; and which were the basis for the G20 “Human-centred AI principles”.¹⁵ To cite just one of the OECD principles:¹⁶

*“AI actors should respect the rule of law, human rights and democratic values. These include freedom, dignity, autonomy, privacy, data protection, non-discrimination, **equality, diversity, fairness, social justice, and internationally recognised labour rights.**”*

- My question is: would you say that autonomous weapons without meaningful human control or those that target humans, respect and are aligned with this principle?
- And if not, why would we consider it acceptable for AI to decide on the right to life, but not on labour rights?
- Fourthly, the **EU Commission’s proposal for new rules and actions for excellence and trust in AI**, of which I would like to highlight three elements:¹⁷
 - i. It takes the courageous step of proposing the **prohibition of practices** including *“all those AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights”*.¹⁸ If right to life is a fundamental right – and it is- why shouldn’t it be prohibited to use autonomous systems that would take it away?
 - ii. The proposal mentions that applications of AI for job recruitment or in the judiciary, among others, should be considered **high-risk and need to be clearly regulated**. If we see these applications as high-risk, why should we not regulate the use of AI in automating violence?
 - iii. The proposed EU regulation differentiates between uses of AI that create (i) **an unacceptable risk**, (ii) **a high risk**, and (iii) **low or minimal**

risk. In other words, it's focus on an assessment of the risks and outcomes of the applications of AI, not on trying to regulate technology itself. This aims to make it future-proof and to avoid fruitless technical discussions. I highlight this because it's a similar approach of the Campaign's proposed normative and operational framework presented by Richard;¹⁹ and really, an emerging trend that we should keep an eye on.

5. Finally, I would like to refer to the **Montréal Declaration for Responsible Development of Artificial Intelligence** signed by 187 organizations, which says – *“The first danger of artificial intelligence development consists in giving the illusion that we can master the future through calculations. Reducing society to a series of numbers and ruling it through algorithmic procedures still drives human ambitions; but when it comes to human affairs, numbers cannot determine what has moral value, nor what is socially desirable”*. It also says, in its principle 9.3: *“The decision to kill must always be made by human beings, responsibility for this decision must not be transferred to an artificial intelligence systems”*.²⁰

There are also regional references such as the *Charter of Ethics on Emerging Technologies in the Arab Region*,²¹ statements from Asian countries on the ethics of AI focusing in promoting harmony and inclusion,²² the recent initiative by Costa Rica called *“For a country free of autonomous weapons: technology for peace”*²³, as well as encouraging news from New Zealand on this specific topic just a few days ago.²⁴

In conclusion, I would like to respectfully call on all of you to reflect on these elements; and examine if your positions on autonomous weapons are coherent with existing commitments in other frameworks.

We believe there is an ethical imperative to act now, because most of us agree that **we as humanity should NOT delegate the selection and engagement of targets to a machine. This is the main issue.** There are other legal, security, technical, operational, and humanitarian arguments; and we are certainly happy to discuss them. But the main point here is that we as society do not want a machine to decide over life and death, based on sensors and algorithms.



Author:

Wanda Muñoz. Victim Assistance & Inclusion Specialist, SEHLAC.
Member of the Global Partnership of Artificial Intelligence.

The author is thankful for the review of the first draft of this statement by **Camilo Serna** (CCCM/SEHLAC Colombia), **Pia Devoto** (APP, SEHLAC Argentina) and **Daan Kayser** (PAX Netherlands).

Contact:



desarme@wandamunoz.com



@SehlaOrg @StopKillerRobots @_WandaMunoz

www.sehla.org

www.stopkillerrobots.org

Notes

¹ See the Future of Life Institute's *Autonomous Weapons: An Open Letter from AI & Robotics Researchers* which has been signed by over 4,500 AI/Robotics researchers and 26,000 other endorses, including Stephen Hawking, Stuart Russell, Francesca Rossi, Toby Walsh, Martha Polack, Félix Castro, Raúl Monroy, Noel Sharkey, Raja Chatila, Mustafa Suleyman...
<https://futureoflife.org/open-letter-autonomous-weapons/>

² See *Machines Capable of Taking Lives Without Human Involvement are Unacceptable*, Secretary General tells experts on autonomous weapons systems.
<https://www.un.org/press/en/2019/sqsm19512.doc.htm> or *Autonomous weapons that kill must be banned, insists UN chief.* <https://news.un.org/en/story/2019/03/1035381>

³ For more information, see: Human Rights Watch. *Stopping Killer Robots: Countries Positions on Banning Killer Robots and Maintaining Meaningful Human Control* (2020).
https://www.hrw.org/sites/default/files/media_2021/04/arms0820_web_1.pdf

Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control
https://www.hrw.org/sites/default/files/media_2021/04/arms0820_web_1.pdf

⁴ For more on the humanitarian concerns raised by autonomous weapons, see ICRC's *Position on Autonomous Weapons Systems*, particularly section 3.2: *Addressing concerns raised by the use of autonomous weapon systems against persons*: <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>

⁵ Para un análisis del impacto de las armas autónomas en las poblaciones más marginadas desde una perspectiva de derechos humanos en América Latina, consultar: Los riesgos de las armas autónomas. Una perspectiva interseccional latinoamericana. <https://bit.ly/3fS76nb> Short video by SEHLAC/IEP Chile: <https://www.youtube.com/watch?v=exPed6m-luQ>

⁶ To see just some examples of these issues:

1. Access to housing: Scheneider, Valérie. *Locked out by big data: how big data, algorithms and machine learning may undermine housing justice*. http://blogs.law.columbia.edu/hrlr/files/2020/11/251_Schneider.pdf
2. Access to employment. *Auditing employment algorithms for discrimination*. <https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/>
3. Access to justice. *Technical Flaws of Pretrial Risk Assessments raise Grave Concerns*. <https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns>
4. Access to health: *Millions of black people affected by racial bias in health-care algorithms*. <https://www.nature.com/articles/d41586-019-03228-6>
5. Access to COVID vaccinations: *This is the Stanford vaccine algorithm that left out frontline doctors. The University hospital blamed a "very complex algorithm" for its unequal vaccine distribution plan* <https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/>

⁷ *Amazon extends moratorium on police use of facial recognition software*: <https://www.reuters.com/technology/exclusive-amazon-extends-moratorium-police-use-facial-recognition-software-2021-05-18/>

⁸ *IBM abandons facial recognition products, condemns racially biased surveillance*: <https://www.npr.org/2020/06/09/873298837/ibm-abandons-facial-recognition-products-condemns-racially-biased-surveillance>

⁹ *Microsoft won't sell police its facial-recognition technology, following similar move by Amazon and IBM* <https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/>
And on the difference between Amazon's and IBM and Microsoft's positions: <https://slate.com/technology/2020/06/ibm-microsoft-amazon-facial-recognition-technology.html>

¹⁰ Buolamwini, J., Gebru, T. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Available at: <https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/> News report on the topic: *Study finds gender and skin-type bias in commercial artificial-intelligence systems*. <https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>
On AI, gender and racial bias, see also "Coded Bias" by the Algorithmic Justice League in Netflix. Trailer available at: <https://www.youtube.com/watch?v=jZI55PsfZJQ> "The people who own the code deploy it on other people, and there is no accountability" – Joy B.

¹¹ One example is the the *Global Partnership of Artificial Intelligence (GPAI)* launched in 2020 at the initiative of the G7: www.gpai.ai GPAI is a "multi-stakeholder initiative which aims to bridge the gap

between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities". Of particular interest for the discussion on autonomous weapons may be the Working Group on Responsible Use of AI and its members, which can be consulted at: <https://www.gpai.ai/projects/responsible-ai/>

GPai's 15 founding members are Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States and the European Union. They were joined by Brazil, the Netherlands, Poland and Spain in December 2020.

¹² Harvard's University Berkman Klein Center has mapped consensus in ethical and rights-based approaches to principles for Artificial Intelligence. In this normative core, the Center has identified *Human Rights Law, Human Control of Technology and Promotion of Human Values* as three of eight thematic trends in artificial intelligence norms, based on analysis of 36 AI Principles documents from around the world. See the report and data visualization: <https://cyber.harvard.edu/publication/2020/principled-ai>

¹³ Access to the current draft of the UNESCO Recommendations by the Ad Hoc Expert Group: <https://unesdoc.unesco.org/ark:/48223/pf0000373434> The last round of negotiations was held in April 2021.

For more information on the process of elaboration of the recommendations led by UNESCO: see <https://en.unesco.org/artificial-intelligence/ethics> According to this website: "***We need international and national policies and regulatory frameworks to ensure that these emerging technologies benefit humanity as a whole. We need a human-centred AI. AI must be for the greater interest of the people, not the other way around.***"

¹⁴ Refers to the "473 Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa - ACHPR/Res. 473 (EXT.OS/ XXXI) 2021. Available in English and French at <https://www.achpr.org/sessions/resolutions?id=504>

¹⁵ The G20 AI principles, which draw from the OECD AI Principles, can be found at the Annex of the G20 Ministerial Statement on Trade and Digital Economy, from the G20 meeting in Japan in 2019: <https://www.mofa.go.jp/files/000486596.pdf>

¹⁶ An overview of the OECD AI Principles can be consulted at: <https://www.oecd.ai/ai-principles> and in pdf format at: <https://bit.ly/2RKnkH9> The official text of the "*Recommendation of the Council on Artificial Intelligence*" can be consulted at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

¹⁷ For an overview, see this press release: "*Europe fit for the digital age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence*": https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682 As mentioned in the presentation, they follow a risk-based approach and clearly state: "**Unacceptable risk: AI systems considered a clear threat to the safety, livelihoods and rights of people will be banned**". The full "Proposal for a Regulation laying down harmonized rules on artificial intelligence" and related annexes can be downloaded at: <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

¹⁸ See the full text of the proposed regulation, pp. 12 Article 5.2.2 *Prohibited Artificial Intelligence Practices (Title II)*

¹⁹ Campaign to Stop Killer Robots' *Advisory Note to the High Contracting Parties to the Convention on Conventional Weapons (CCW)* May 2021. Available at: <https://drive.google.com/file/d/1TrjOf0K1Gc9BGVkcfoaNCY6Vw5wc5eAB/view>

²⁰ Website of the Montréal Declaration: <https://www.montrealdeclaration-responsibleai.com> Access to the text of the Declaration: <https://www.montrealdeclaration-responsibleai.com/the-declaration> Comment on the Declaration by Canadian AI expert Yoshua Bengio: <https://theconversation.com/the-montreal-declaration-why-we-must-develop-ai-responsibly-108154>

²¹ This Charter can be downloaded at: http://www.umi.ac.ma/wp-content/uploads/2020/01/Charte-Ethique-des-Sciences-et-Technologies-UNESCO-V.Eng_.pdf

²² See for example *Experts from Central Asia debate the development of artificial intelligence in the region*, which says “*Debating the openness of AI development, participants focused on how to foster clarity and transparency of algorithms, since the “blackboxes” – the lack of transparency- in automatized decision-making processes pose major challenges occurring alongside the development of AI...*” Available at: <https://bit.ly/3i22VrM> Also: *ROK and UNESCO co-organize virtual Asia Pacific Consultation on UNESCO Recommendation on the Ethics of Artificial Intelligence*. In this framework, the Republic of Korea stated its “...strong commitment to objectives such as “living in harmony” and “leaving no one behind” which are addressed in the draft text of the UNESCO Recommendation, and furthermore reiterated Korea’s willingness to strengthen cooperation, based on value such as solidarity and inclusion, with UNESCO member countries in the Asia-Pacific and beyond”. Available at: <https://bit.ly/3fuVMOP>

²³ See: “Costa Rica lanza Campaña para la promoción de la tecnología al servicio de la paz” <https://www.micit.go.cr/noticias/costa-rica-lanza-campana-la-promocion-la-tecnologia-al-servicio-la-paz>

²⁴ Hon. Phil Twyford, Minister for Disarmament and Arms Control of New Zealand, stated on May 26th 2021: “*I have committed New Zealand to building an alliance of countries working towards an international and legally-binding instrument prohibiting and regulating unacceptable autonomous weapons systems...*” See: *Shaping the future: Autonomous Weapons Systems* at <https://www.beehive.govt.nz/speech/shaping-future-autonomous-weapons-systems>